

PAGE: 1

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/938,548BDATE: 09/17/98  
TIME: 09:12:00

INPUT SET: S28690.raw

This Raw Listing contains the General  
Information Section and up to the first 5 pages

ENTERED

## SEQUENCE LISTING

## (1) General Information

(i) APPLICANT: Yanagisawa, Masashi  
Bergsma, Derk  
Wilson, Shelagh  
Brooks, David  
Gellai, Miklos

(ii) TITLE OF THE INVENTION: NOVEL LIGANDS OF THE NEUROPEPTIDE  
RECEPTOR HFGAN72

(iii) NUMBER OF SEQUENCES: 21

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SmithKline Beecham Corporation  
(B) STREET: 709 Swedeland Road  
(C) CITY: King of Prussia  
(D) STATE: PA  
(E) COUNTRY: United States of America  
(F) ZIP: 19406

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette  
(B) COMPUTER: IBM Compatible  
(C) OPERATING SYSTEM: DOS  
(D) SOFTWARE: FastSEQ for Windows Version 2.0

## (vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 08/938,548  
(B) FILING DATE: 26-SEPT-1997  
(C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/887,382  
(B) FILING DATE: 2-JUL-1997

## (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/820,519  
(B) FILING DATE: 19-MAR-1997

(A) APPLICATION NUMBER: 60/033,604  
(B) FILING DATE: 17-DEC-1997

## (viii) ATTORNEY/AGENT INFORMATION:

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/938,548BDATE: 09/17/98  
TIME: 09:12:02

INPUT SET: S28690.raw

47 (A) NAME: Elizabeth J. Hecht  
48 (B) REGISTRATION NUMBER: 41,824  
49 (C) REFERENCE/DOCKET NUMBER: ATG50037-2  
50

51 (ix) TELECOMMUNICATION INFORMATION:

52 (A) TELEPHONE: 610-270-5009

53 (B) TELEFAX: 610-270-5090

54 (C) TELEX:  
55

56 (2) INFORMATION FOR SEQ ID NO:1:  
57

58 (i) SEQUENCE CHARACTERISTICS:

59 (A) LENGTH: 1970 base pairs

60 (B) TYPE: nucleic acid

61 (C) STRANDEDNESS: single

62 (D) TOPOLOGY: linear  
63

64 (ii) MOLECULE TYPE: Genomic DNA  
65

66 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:  
67

68	AAAACATAAT	GTGGGTCTCG	CGTCTGCCTC	TCTCCCGCCC	CTAATTAGCA	GCTGCCTCCC	60
69	TCCATATTGT	CCCAGGCCAG	CGCTTCTTTT	GTGCTCCAG	ATTCCCTGGGT	GCAAGGTGGC	120
70	CTCATTAGTG	CCCGGAGACC	GCCCCATCTC	CAGGGAGCAG	ATAGACAGAC	AAGGGGGTGA	180
71	TCAGGGGCAC	AGTGATCCAA	CCCTGGCCTC	TGAACGCCGC	AGCGGCCATT	CCTTGGGCCC	240
72	AGCCTGGAGA	CGGCCCCCCT	GCAGCAGGCT	AATCTTAGAC	TTGCCCTTTGT	CTGGCCTGGG	300
73	TGTGGACGCA	ATGTGCCTGT	CAATTCCCCG	CCACCTCAGA	GCACATATAA	CCCCAGACCC	360
74	CTGGGAGTGG	GTCACAATTG	ACAGCCTCAA	GGTTCCTGGC	TTTTTGAACC	ACCACAGACA	420
75	TCTCCTTTCC	CGGCTACCCC	ACCCTGAGCG	CCAGACACCA	TGAACCTTCC	TTCCACAAAG	480
76	GTAAAGATCC	AGGGATGGAG	GGGTGACTCA	GCCATCCCAG	AGGAAGCAAA	AAGAGTGCCT	540
77	GCTCAGAGGG	CTGGAAGAAA	GGCCAAAGGT	GTCTCCACTC	TTGGTCTTTT	CCTGGGTGTG	600
78	CTCTGAGGCA	GGAGCACCTG	CCTTGGCTCA	CATTGGGTTG	GGTGCCTGTT	TGCTAAGAGC	660
79	CTGTGTTTGC	TGAGCTCATA	TGTGTCAGGT	GCTCCGTTTG	CACCTGTCAT	CTCTTGTCAT	720
80	CCTCCCAACA	GCCTTGACAG	GTAAGAAATTA	TTTCTAGTAT	ACCCAGTTTA	CAGGTAAGGG	780
81	AGCTGTGCCC	TCTGAAAGGG	CAGGAAACTG	GTTCAAAGCA	ACGGAGTTCA	GTCACCTCTG	840
82	CAAGGGGGCA	GGCAGATGAG	AGAGCATTCT	GGAGTCTTGC	TAGTTCCTGA	TTTCCATGTG	900
83	TTTCCCTGCT	GTGGAGAGGA	AGTTGGGGGG	ACTCAGTAGG	GCCCCGGTTT	TTCCCAAGTT	960
84	TACAACCTTCT	GCTGCAGACA	GACACTCCTG	TTTTTCAGGTG	GAGTGGCAAG	TGCCCTAGTG	1020
85	GTGGCAACAG	TGGCCTAAGT	CTCCAGAGAA	AAGGGGGATT	CACTCTGCCC	AGGGGGTCTC	1080
86	AAAAGGCTTC	CTGTGGGAGA	TGCTCTGCTG	GGTCTTGAAG	GAGGAGCAGG	GAAAGTAGGC	1140
87	CGATACCAGC	AAGGGCGCAA	AGCAAGGAGA	ACTAAGTGAC	AGCCAGAAAG	GAGTGCAGGC	1200
88	TTGGAGGGGG	CGCGGAGCCA	GAGGGGACAG	TCCTGTGCGT	GGGAGCTGGT	GGCGGGCGCC	1260
89	GTGGGAAGAC	CCCCCCAGCG	CCCTGTCTCC	GTCTCCCTAG	GTCTCCTGGG	CCGCCGTGAC	1320
90	GCTACTGCTG	CTGCTGCTGC	TGCTGCCGCC	CGCGCTGTTG	TCGTCCGGGG	CGGCTGCACA	1380
91	GCCCCTGCCC	GACTGCTGTC	GTCAAAAGAC	TTGCTCTTGC	CGCCTCTACG	AGCTGCTGCA	1440
92	CGGCGCGGGC	AATCACGCGG	CCGGCATCCT	CACGCTGGGC	AAGCGGAGGT	CCGGGCCCCC	1500
93	GGGCCTCCAG	GGTCGGCTGC	AGCGCCTCCT	GCAGGCCAGC	GGCAACCACG	CCGCGGGCAT	1560
94	CCTGACCATG	GGCCGCCGCG	CAGGCGCAGA	GCCAGCGCCG	CGCCCCTGCC	TCGGGCGCCG	1620
95	CTGTTCCGCC	CCGGCCGCCG	CCTCCGTCGC	GCCCGGAGGA	CAGTCCGGGA	TCTGAGTCGT	1680
96	TCTTCGGGCC	CTGTCTTGGC	CCAGGCCCTCT	GCCCTCTGCC	CACCCAGCGT	CAGCCCCCAG	1740
97	AAAAAAGGCA	ATAAAGACGA	GTCTCCATTC	GTGTGACTGG	TCTCTGTTCC	TGTGCGGTCC	1800
98	CGTCCTGCCC	ATCCGGGGTG	GCAAAGCGTC	TTGCGGAGGA	CAGCTGGGCC	TGGAAGCCCC	1860
99	GCTGTCGGGC	ACCAGCCTTA	GCTTTTGCCT	GGTTGAATCG	GAAACACTCT	TGTTTGGGGA	1920

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/938,548BDATE: 09/17/98  
TIME: 09:12:04

INPUT SET: S28690.raw

1970

100 GTTCCCAGTG CAAGGCCCTG GGGCACAGAG AGAACTGCAC AGGTGCATGC

101

102 (2) INFORMATION FOR SEQ ID NO:2:

103

104 (i) SEQUENCE CHARACTERISTICS:

105 (A) LENGTH: 131 amino acids

106 (B) TYPE: amino acid

107 (C) STRANDEDNESS: single

108 (D) TOPOLOGY: linear

109

110 (ii) MOLECULE TYPE: protein

111

112 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

113

114 Met Asn Leu Pro Ser Thr Lys Val Ser Trp Ala Ala Val Thr Leu Leu

115 1 5 10 15

116 Leu Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Ser Gly Ala Ala

117 20 25 30

118 Ala Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg

119 35 40 45

120

121 Leu Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu

122 50 55 60

123 Thr Leu Gly Lys Arg Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu

124 65 70 75 80

125 Gln Arg Leu Leu Gln Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr

126 85 90 95

127 Met Gly Arg Arg Ala Gly Ala Glu Pro Ala Pro Arg Pro Cys Leu Gly

128 100 105 110

129 Arg Arg Cys Ser Ala Pro Ala Ala Ser Val Ala Pro Gly Gly Gln

130 115 120 125

131 Ser Gly Ile

132 130

133

134 (2) INFORMATION FOR SEQ ID NO:3:

135

136 (i) SEQUENCE CHARACTERISTICS:

137 (A) LENGTH: 33 amino acids

138 (B) TYPE: amino acid

139 (C) STRANDEDNESS: single

140 (D) TOPOLOGY: linear

141

142 (ii) MOLECULE TYPE: protein

143

144 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

145

146 Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu

147 1 5 10 15

148 Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr

149 20 25 30

150 Leu

151

152 (2) INFORMATION FOR SEQ ID NO:4:

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/938,548BDATE: 09/17/98  
TIME: 09:12:06

INPUT SET: S28690.raw

153

154

(i) SEQUENCE CHARACTERISTICS:

155

(A) LENGTH: 28 amino acids

156

(B) TYPE: amino acid

157

(C) STRANDEDNESS: single

158

(D) TOPOLOGY: linear

159

160

(ii) MOLECULE TYPE: protein

161

162

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

163

Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln Arg Leu Leu Gln

164

1

5

10

15

165

Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr Met

166

20

25

167

168

169

(2) INFORMATION FOR SEQ ID NO:5:

170

171

(i) SEQUENCE CHARACTERISTICS:

172

(A) LENGTH: 585 base pairs

173

(B) TYPE: nucleic acid

174

(C) STRANDEDNESS: single

175

(D) TOPOLOGY: linear

176

177

(ii) MOLECULE TYPE: cDNA

178

179

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

180

181

GGCTCGGCGG CCTCAGACTC CTTGGGTATT TGGACCACTG CACCGAAGAT ACCATCTCTC 60

182

CGGATTGCCT CTCCCTGAGC TCCAGACACC ATGAACCTTC CTTCTACAAA GGTTCCCTGG 120

183

GCCGCCGTGA CGCTGCTGCT GCTGCTACTG CTGCCGCCGG CGCTGCTGTC GCTTGGGGTG 180

184

GACGCGCAGC CTCTGCCCGA CTGCTGTCTG CAGAAGACGT GTTCCTGCCG TCTCTACGAA 240

185

CTGTTGCACG GAGCTGGCAA CCACGCCGCG GGCATCCTCA CTCTGGGAAA GCGGCGACCT 300

186

GGACCCCCAG GCCTCCAAGG ACGGCTGCAG CGCCTCCTTC AGGCCAACGG TAACCACGCA 360

187

GCTGGCATCC TGACCATGGG CCGCCGCGCA GGCGCAGAGC TAGAGCCATA TCCCTGCCCT 420

188

GGTCGCCGCT GTCCGACTGC AACCGCCACC GCTTTAGCGC CCCGGGGCGG ATCCAGAGTC 480

189

TGAACCCGTC TTCTATCCCT GTCCTAGTCC TAACTTTCCC CTCTCCTCGC CGGTCCCTAG 540

190

GCAATAAAGA CGTTTCTCTG CTAATAAAAAA AAAAAAAAAA AAAAA 585

191

192

(2) INFORMATION FOR SEQ ID NO:6:

193

194

(i) SEQUENCE CHARACTERISTICS:

195

(A) LENGTH: 130 amino acids

196

(B) TYPE: amino acid

197

(C) STRANDEDNESS: single

198

(D) TOPOLOGY: linear

199

200

(ii) MOLECULE TYPE: protein

201

202

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

203

204

Met Asn Leu Pro Ser Thr Lys Val Pro Trp Ala Ala Val Thr Leu Leu

205

1

5

10

15

# RAW SEQUENCE LISTING PATENT APPLICATION US/08/938,548B

DATE: 09/17/98  
TIME: 09:12:08

INPUT SET: S28690.raw

```

206  Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Leu Gly Val Asp Ala
207              20              25              30
208  Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
209              35              40              45
210  Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
211              50              55              60
212  Leu Gly Lys Arg Arg Pro Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln
213              65              70              75              80
214  Arg Leu Leu Gln Ala Asn Gly Asn His Ala Ala Gly Ile Leu Thr Met
215              85              90              95
216  Gly Arg Arg Ala Gly Ala Glu Leu Glu Pro Tyr Pro Cys Pro Gly Arg
217              100             105             110
218  Arg Cys Pro Thr Ala Thr Ala Thr Ala Leu Ala Pro Arg Gly Gly Ser
219              115             120             125
220  Arg Val
221      130
222

```

## (2) INFORMATION FOR SEQ ID NO:7:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: protein

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```

235  Met Asn Leu Pro Ser Thr Lys Val Pro Trp Ala Ala Val Thr Leu Leu
236      1              5              10              15
237  Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Leu Gly Val Asp Ala
238              20              25              30
239

```

## (2) INFORMATION FOR SEQ ID NO:8:

### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

### (ii) MOLECULE TYPE: protein

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

252  Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
253      1              5              10              15
254  Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
255              20              25              30
256  Leu
257
258

```

PAGE: 1

**SEQUENCE VERIFICATION REPORT**  
**PATENT APPLICATION US/08/938,548B**

DATE: 09/17/98  
TIME: 09:12:10

*INPUT SET: S28690.raw*

Line

Error

Original Text